CLAIMS

We claim:

1. A method of producing binary sequence tags from nucleic acid fragments in a nucleic acid sample, the method comprising

(a) incubating a nucleic acid sample with one or more first nucleic acid cleaving reagents to produce nucleic acid fragments,

(b) mixing one or more offset adaptors with the nucleic acid sample and covalently coupling the offset adaptors to the nucleic acid fragments,

(c) incubating the nucleic acid sample with one or more second nucleic acid cleaving reagents to produce nucleic acid fragments with sticky ends, wherein the second nucleic acid cleaving reagents cleave at a site offset from their recognition sequence, wherein each offset adaptor has a recognition sequence for at least one of the second nucleic acid cleaving reagents,

(d) mixing one or more adaptor-indexers with the nucleic acid sample and covalently coupling the adaptor-indexers to the nucleic acid fragments, wherein each adaptor-indexer has a different sticky end, wherein each sticky end of the adaptor-indexers is compatible with a sticky end generated by the second nucleic acid cleaving reagents,

wherein the nucleic acid fragments to which offset adaptors and adaptor-indexers have been coupled are binary sequence tags.

2. The method of claim 1 wherein the binary sequence tags are amplified, detected, identified, sequenced, cataloged, or a combination.

3. The method of claim 2 wherein the binary sequence tags are detected, wherein detection comprises determining, directly or indirectly, the presence, amount, presence and amount, or absence of one or more binary sequence tags.

4. The method of claim 1 wherein the ~~offset adaptors~~ and the adaptor-indexers are covalently coupled to the nucleic acid fragments by ligation.

5. The method of claim 1 wherein the first and second nucleic acid cleaving reagents are restriction enzymes.

6. The method of claim 5 wherein the recognition sequence of the restriction enzymes are from four to thirty nucleotides in length.

7. The method of claim 5 wherein the recognition sequence of the restriction enzymes are from four to ten nucleotides in length.

8. The method of claim 5 wherein the recognition sequence of the restriction enzymes are from four to eight nucleotides in length.

9. The method of claim 5 wherein the first restriction enzymes have a four base recognition sequence, wherein one second restriction enzyme is used, and wherein the second restriction enzyme is a Type IIS restriction enzyme that cleaves at a site different from its recognition sequence.

10. The method of claim 1 wherein binary sequence tags are amplified.

11. The method of claim 10 wherein the binary sequence tags are labeled during amplification.

12. The method of claim 10 wherein binary sequence tags are amplified by the polymerase chain reaction.

13. The method of claim 1 further comprising

hybridizing the binary sequence tags to detector probes.

14. The method of claim 13 wherein the detector probes are coupled to beads.

15. The method of claim 14 wherein the beads contain a label.

16. The method of claim 15 wherein the label is a molecular barcode.

17. The method of claim 15 wherein the label is a mass label.

18. The method of claim 15 further comprising

sorting or separating the binary sequence tags via the labels.

19. The method of claim 14 further comprising

sorting or separating the binary sequence tags via the beads.

20. The method of claim 13 wherein the detector probes are immobilized on a substrate in an array.

21. The method of claim 13 wherein the detector probes contain a label.

22. The method of claim 21 further comprising

sorting or separating the binary sequence tags via the labels.

23. The method of claim 1 further comprising

detecting the binary sequence tags by mass spectroscopy.

24. The method of claim 23 wherein the binary sequence tags are detected by matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy.

25. The method of claim 23 wherein, prior to, or during detection the binary sequence tags are fragmented by collisionally induced dissociation.

26. The method of claim 23 wherein one or both strands of the binary sequence tags are cleaved or partially degraded prior to detection.

27. The method of claim 26 wherein the binary sequence tags are partially degraded using Texaphyrin technology.

28. The method of claim 23 wherein the binary sequence tags are hybridized to detector probes prior to detection.

29. The method of claim 28 wherein the detector probes are immobilized on a substrate in an array.

30. The method of claim 1 further comprising

sequencing all or a portion of one or more of the binary sequence tags.

31. The method of claim 30 wherein the portion of the binary sequence tag that does not correspond to sequence in either the offset adaptor or the adaptor-indexer is sequenced.

32. The method of claim 30 further comprising, prior to sequencing

diluting each nucleic acid sample and dividing into aliquots containing, on average, one binary sequence tag molecule per aliquot, and

amplifying the binary sequence tag in each aliquot.

33. The method of claim 1 wherein each adaptor-indexer has a double-stranded portion, wherein the double-stranded portion of each adaptor-indexers is different.

34. The method of claim 1 wherein each adaptor-indexer has a double-stranded portion, wherein the double-stranded portion of each adaptor-indexers is the same.

35. The method of claim 1 wherein each adaptor-indexer has a double-stranded portion, wherein the double-stranded portions of at least two of adaptor-indexers are different.

36. The method of claim 1 wherein each adaptor-indexer has a double-stranded portion, wherein the double-stranded portions of at least two of adaptor-indexers are the same.

37. The method of claim 1 wherein each adaptor-indexer contains a label.

38. The method of claim 37 further comprising

sorting or separating the binary sequence tags via the labels.

39. The method of claim 1 further comprising

separating the strands of the binary sequence tags.

40. The method of claim 39 wherein each binary sequence tag comprises a top strand and a bottom strand,

wherein the top strands have offset adaptor sequence at their 5' ends, wherein the bottom strands have adaptor-indexer sequence at their 5' ends,

wherein the top strands of the binary sequence tags are separated from the bottom strands of the binary sequence tags.

41. The method of claim 40 further comprising

hybridizing one or more ligator-detectors with either the top strands of the binary sequence tags or the bottom strands of the binary sequence tags, wherein each ligator-detector comprises sequence matching or complementary to all or part of sequence including, and adjacent to, the sticky end of at least one of the adaptor-indexers,

hybridizing the nucleic acid sample with a detector array comprising one or more probes and covalently coupling the ligator-detectors to the probes, wherein each probe has a different sequence, and

detecting, directly or indirectly, coupling of ligator-detectors to the detector array probes.

42. The method of claim 1 further comprising

hybridizing one or more ligator-detectors with the binary sequence tags, wherein each ligator-detector comprises sequence matching or complementary to all or part of sequence including, and adjacent to, the sticky end of at least one of the adaptor-indexers,

hybridizing the nucleic acid sample with a detector array comprising one or more probes and covalently coupling the ligator-detectors to the probes, wherein each probe has a different sequence, and

detecting, directly or indirectly, coupling of ligator-detectors to the detector array probes.

43. The method of claim 42 wherein coupling of the ligator-detector is detected by rolling circle replication of an amplification target circle wherein replication is primed by the ligator-detector.

44. The method of claim 43 wherein uncoupled ligator-detectors do not prime rolling circle replication of an amplification target circle.

45. The method of claim 42 wherein the nucleic acid cleaving reagents generate sticky ends having N different sequences, and wherein the sample is divided into N index samples.

46. The method of claim 42 wherein the detector array probes are all of the same length.

47. The method of claim 46 wherein the detector array probes are six, seven, or eight nucleotides long.

48. The method of claim 42 wherein the detector array probes all have similar hybrid stability.

49. The method of claim 42 wherein each offset adaptor, adaptor-indexer, ligator-detector, or detector array probe contains a label, wherein coupling of the ligator-detectors to the probes is detected via the label.

50. The method of claim 1 wherein the collection of binary sequence tags produced from the nucleic acid sample constitutes a catalog of binary sequence tags for the nucleic acid sample.

51. The method of claim 1 further comprising

performing steps (a) through (d) on a plurality of nucleic acid samples.

52. The method of claim 51 further comprising

performing steps (a) through (d) on a control nucleic acid sample,

identifying differences between the collection of binary sequence tags produced from the nucleic acid samples and the control nucleic acid sample.

53. The method of claim 52 wherein the differences are differences in the presence, amount, presence and amount, or absence of binary sequence tags produced from the nucleic acid samples and the control nucleic acid sample.

54. The method of claim 51 wherein the steps (a) through (d) are performed on a control nucleic acid sample and a tester nucleic acid sample,

wherein the tester nucleic acid sample, or the source of the tester nucleic acid sample, is treated, prior to step (a), so as to destroy, disrupt or eliminate one or more nucleic acid molecules in the tester nucleic acid sample,

wherein the binary sequence tags corresponding to the destroyed, disrupted, or eliminated nucleic acid molecules will be produced from the control nucleic acid sample but not the tester nucleic acid sample.

55. The method of claim 54 wherein the tester nucleic acid sample is treated so as to destroy, disrupt or eliminate one or more nucleic acid molecules in the tester nucleic acid sample.

56. The method of claim 55 wherein the tester and control nucleic acid samples are samples of messenger RNA, the method further comprising, prior to step (a)

reverse transcribing the messenger RNA molecules to produce first cDNA strands of the messenger RNA molecules,

destroying or disrupting one or more first cDNA strands by sequence-specific cleavage in the tester nucleic acid sample but not the control nucleic acid sample,

synthesizing second cDNA strands from the first DNA strands.

57. The method of claim 55 wherein the tester and control nucleic acid samples are samples of messenger RNA, the method further comprising, prior to step (a)

reverse transcribing the messenger RNA molecules to produce first and second cDNA strands of the messenger RNA molecules,

destroying or disrupting one or more second cDNA strands by sequence-specific cleavage in the tester nucleic acid sample but not the control nucleic acid sample.

58. The method of claim 54 wherein the source of the tester nucleic acid sample is treated so as to destroy, disrupt or eliminate one or more nucleic acid molecules in the tester nucleic acid sample.

59. The method of claim 58 wherein the treatment of the source is accomplished by exposing cells from which the tester sample will be derived with a compound, composition, or condition that ~~will~~ reduce or eliminate expression of one or more genes.

60. The method of claim 54 further comprising identifying differences in the binary sequence tags produced from the control nucleic acid sample and tester nucleic acid sample.

61. The method of claim 51 further comprising

identifying differences between the collection of binary sequence tags produced from the nucleic acid samples.

62. The method of claim 61 further comprising identifying or preparing nucleic acid fragments corresponding the binary sequence tags produced from one or more of the nucleic acid samples but not from one or more of the other nucleic acid samples.

63. The method of claim 1 further comprising identifying or preparing nucleic acid fragments corresponding the binary sequence tags produced from the nucleic acid sample.

64. The method of claim 63 further comprising using the prepared nucleic acid fragments as probes for analyzing a different nucleic acid sample.

65. The method of claim 64 wherein analysis of the different nucleic acid sample involves detection, quantitation, identification, comparison, screening,

sequencing, culling, destruction, sorting, capturing, or a combination, of nucleic acid molecules in the different nucleic acid sample.

66. The method of claim 1 wherein a single type of offset adaptor is used.

67. The method of claim 1 wherein one second nucleic acid cleaving reagent is used, wherein the second nucleic acid cleaving reagent is a Type IIS restriction enzyme.

68. The method of claim 1 wherein at least one of the first nucleic acid cleaving reagents is sensitive to modification of its recognition site.

69. The method of claim 68 wherein the first and second nucleic acid cleaving reagents are restriction enzymes.

70. The method of claim 69 wherein the modification to the recognition site is methylation, alkylation, dimerization, derivatization, depurination, or ADP-ribosylation.

71. The method of claim 69 wherein the modification is present in the nucleic acid fragments when isolated or is introduced to the nucleic acid fragments after isolation.

72. The method of claim 69 further comprising, following coupling of the adaptor-indexers to the nucleic acid fragments,

amplifying the nucleic acid fragments in the index samples to which offset adaptors and adaptor-indexers have been coupled.

73. The method of claim 72 further comprising determining the sequence of a portion of at least one of the nucleic acid fragments in the nucleic acid sample.

74. The method of claim 69 further comprising, following coupling of the offset adaptors to the nucleic acid fragments,

separating nucleic acid fragments coupled to offset adaptors from nucleic acid fragments not coupled to offset adaptors, wherein only nucleic acid fragments coupled to offset adaptors are used in step (c).

75. The method of claim 69 wherein at least one of the first restriction enzymes (1) is insensitive to modification of its recognition site and (2) has the same

recognition site as the first restriction enzyme that is sensitive to modification of its recognition site,

the method further comprising, prior to digestion with the first restriction enzymes,

dividing each index sample into a set of two or more of index samples,

wherein each secondary index sample in each set of index samples is digested with a different first restriction enzyme,

wherein steps (a) through (d) are performed with each of the secondary index samples.

76. The method of claim 75 further comprising,

comparing the pattern of the presence or absence of binary sequence tags made using the first restriction enzyme that is sensitive to modification of its recognition site with the pattern of the presence or absence of binary sequence tags made using the first restriction enzyme that is insensitive to modification of its recognition site and that has the same recognition site as the first restriction enzyme that is sensitive to modification of its recognition site, ,

wherein differences in the patterns indicate modification of nucleic acids in the nucleic acid sample.

77. The method of claim 69 wherein the pattern of the presence, amount, presence and amount, or absence of binary sequence tags constitutes a catalog of nucleic acid fragments in the nucleic acid sample.

78. The method of claim 77 further comprising preparing a second catalog of nucleic acid fragments in a second nucleic acid sample and comparing the first catalog and second catalog, wherein differences in the first and second catalogs indicate differences in modification of the first and second nucleic acid samples.

79. The method of claim 77 wherein the second nucleic acid sample is a sample from the same type of cells as the first nucleic acid sample except that the cells from which the first nucleic acid sample is derived are modification-deficient relative to the cells from which the second nucleic acid sample is derived.

80. The method of claim 77 wherein the second nucleic acid sample is a sample from a different type of cells than the first nucleic acid sample, and wherein the cells from which the first nucleic acid sample is derived are modification-deficient relative to the cells from which the second nucleic acid sample is derived.

81. The method of claim 1 wherein the nucleic acid sample is a cDNA sample, wherein the cDNA is synthesized in the presence of one or more dideoxy nucleoside triphosphates,

the method further comprising

identifying correlated binary sequence tags.

82. The method of claim 81 wherein the correlated binary sequence tags are identified by

calculating the expression levels and the expression ratios of a plurality of the binary sequence tags,

identifying binary sequence tags that have similar expression ratios,

grouping the binary sequence tags that have similar expression ratios in binary pairs, to generate a list of correlated binary sequence tags,

ordering the binary pairs in the list of correlated binary sequence tags according to expression level of the binary sequence tags, and

fitting the values of the expression levels to a standard curve based on an inverse exponential function.

83. The method of claim 82 further comprising

identifying correlated binary sequence tags using a different a cDNA sample, wherein the cDNA is synthesized in the presence of a different concentration of the one or more dideoxy nucleoside triphosphates or a different set of one or more dideoxy nucleoside triphosphates,

correlating changes in expression of each list of correlated binary sequence tags with the predicted slope changes predicted by the inverse exponential function corresponding to each of the dideoxy terminator levels, and

predicting the order and position within the cDNA of each list of binary tags.

84. The method of claim 81 wherein the cDNA is synthesized using one or more anchored primers.

85. The method of claim 1 further comprising

performing steps (a) through (d) on a second nucleic acid sample.

86. The method of claim 85 wherein the second nucleic acid sample is a sample from the same type of organism as the first nucleic acid sample.

87. The method of claim 85 wherein the second nucleic acid sample is a sample from the same type of tissue as the first nucleic acid sample.

88. The method of claim 85 wherein the second nucleic acid sample is a sample from the same organism as the first nucleic acid sample.

89. The method of claim 88 wherein the second nucleic acid sample is obtained at a different time than the first nucleic acid sample.

90. The method of claim 85 wherein the second nucleic acid sample is a sample from a different organism than the first nucleic acid sample.

91. The method of claim 85 wherein the second nucleic acid sample is a sample from a different type of tissue than the first nucleic acid sample.

92. The method of claim 85 wherein the second nucleic acid sample is a sample from a different species of organism than the first nucleic acid sample.

93. The method of claim 85 wherein the second nucleic acid sample is a sample from a different strain of organism than the first nucleic acid sample.

94. The method of claim 85 wherein the second nucleic acid sample is a sample from a different cellular compartment than the first nucleic acid sample.

95. The method of claim 1 further comprising,

dividing the nucleic acid sample into a plurality of index samples.

96. The method of claim 95 wherein a different adaptor-indexer is mixed with each index sample.

97. The method of claim 95 wherein a different second nucleic acid cleaving reagent is mixed with each index sample.

98. The method of claim 95 wherein a different offset adaptor is mixed with each index sample.

99. The method of claim 95 wherein a different first nucleic acid cleaving reagent is mixed with each index sample.

100. The method of claim 95 wherein the nucleic acid sample is divided into a plurality of index samples prior to step (d).

101. The method of claim 95 wherein the nucleic acid sample is divided into a plurality of index samples prior to step (c).

102. The method of claim 95 wherein the nucleic acid sample is divided into a plurality of index samples prior to step (b).

103. The method of claim 95 wherein the nucleic acid sample is divided into a plurality of index samples prior to step (a).

104. The method of claim 95 wherein one or more of the index samples are divided into a plurality of secondary index samples.

105. The method of claim 104 wherein a different adaptor-indexer is mixed with each secondary index sample.

106. The method of claim 104 wherein a different second nucleic acid cleaving reagent is mixed with each secondary index sample.

107. The method of claim 104 wherein a different offset adaptor is mixed with each secondary index sample.

108. The method of claim 104 wherein one or more of the secondary index samples are divided into a plurality of tertiary index samples.

109. The method of claim 108 wherein a different adaptor-indexer is mixed with each tertiary index sample.

110. The method of claim 108 wherein a different second nucleic acid cleaving reagent is mixed with each tertiary index sample.

111. The method of claim 95 wherein the binary sequence tags are amplified prior to dividing the nucleic acid sample into a plurality of index samples.

112. The method of claim 95 wherein the binary sequence tags are amplified following dividing the nucleic acid sample into a plurality of index samples.

113. The method of claim 112 further comprising

prior to amplifying the binary sequence tags, diluting each index sample and dividing into aliquots containing, on average, one binary sequence tag molecule per aliquot, and

following amplifying the binary sequence tags, sequencing all or a portion of one or more of the binary sequence tags are sequenced.

114. The method of claim 1 further comprising, following step (d),

mixing the nucleic acid sample with one or more different hairpin primers, wherein each hairpin primer comprises a different primer sequence, wherein each primer sequence is complementary to all or part of the sequence of at least one of the adaptor-indexers,

incubating the nucleic acid sample under conditions that promote amplification of binary sequence tags, wherein amplified binary sequence tags are formed which have hairpin primer sequences at one or both ends,

incubating the nucleic acid sample under conditions that promote formation of hairpin structures by the hairpin primer sequences at the ends of the amplified binary sequence tags,

hybridizing the nucleic acid sample with a plurality of detector probes and covalently coupling the hairpin structures to the probes, wherein each probe has a different sequence, and

detecting coupling of the amplified fragments to different detector probes.

115. The method of claim 1 wherein the concentration of the binary sequence tags is normalized.

116. The method of claim 115 wherein the concentration of the binary sequence tags is normalized by immobilizing one strand of the binary sequence tags, denaturing the binary sequence tags, renaturing the binary sequence tags for a time greater than the $c_0t_{1/2}$ for abundant binary sequence tags and less than the $c_0t_{1/2}$ for rare binary sequence tags, and collecting the un-renatured binary sequence tags.

Sub A5

117. A method of producing binary sequence tags from nucleic acid fragments in a nucleic acid sample, the method comprising

Sub A3

(a) incubating a nucleic acid sample with one or more first nucleic acid cleaving reagents to produce nucleic acid fragments,

(b) mixing one or more first offset adaptor strands with the nucleic acid sample and covalently coupling the first offset adaptor strands to the nucleic acid fragments, wherein, after coupling, the first offset adaptor strands are fully or partially single-stranded,

(c) treating the nucleic acid sample to result in full or partial complementary sequences hybridized to the first offset adaptor strands,

(d) incubating the nucleic acid sample with one or more second nucleic acid cleaving reagents to produce nucleic acid fragments with sticky ends, wherein the second nucleic acid cleaving reagents cleave at a site offset from their recognition sequence, wherein each first offset adaptor strand has a recognition sequence for at least one of the second nucleic acid cleaving reagents,

(e) mixing one or more adaptor-indexers with the nucleic acid sample and covalently coupling the adaptor-indexers to the nucleic acid fragments, wherein each adaptor-indexer has a different sticky end, wherein each sticky end of the adaptor-indexers is compatible with a sticky end generated by the second nucleic acid cleaving reagents,

wherein the nucleic acid fragments to which offset adaptors and adaptor-indexers have been coupled are binary sequence tags.

118. The method of claim 117 wherein treatment of the nucleic acid sample to result in full or partial complementary sequences hybridized to the first offset adaptor strands is accomplished by hybridizing second offset adaptor strands to the coupled first offset adaptor strands.

119. The method of claim 118 further comprising covalently coupling the second offset adaptor strands to the nucleic acid fragments.

120. The method of claim 117 wherein treatment of the nucleic acid sample to result in full or partial complementary sequences hybridized to the first offset adaptor strands is accomplished by filling in the single-stranded portion of the first offset adaptor strands.

121. A method of producing binary sequence tags from nucleic acid fragments in a nucleic acid sample, the method comprising

(a) incubating a nucleic acid sample with one or more first nucleic acid cleaving reagents to produce nucleic acid fragments,

Sub A5

(b) mixing one or more offset adaptors with the nucleic acid sample and covalently coupling the offset adaptors to the nucleic acid fragments,

(d) incubating the nucleic acid sample with one or more second nucleic acid cleaving reagents to produce nucleic acid fragments with sticky ends, wherein the second nucleic acid cleaving reagents cleave at a site offset from their recognition sequence, wherein each first offset adaptor strand has a recognition sequence for at least one of the second nucleic acid cleaving reagents,

(e) mixing one or more first adaptor-indexer strands with the nucleic acid sample and covalently coupling the first adaptor-indexer strands to the nucleic acid fragments, wherein each first adaptor-indexer strand has a different end sequence, wherein each end sequence of the first adaptor-indexer strands is compatible with a sticky end generated by the second nucleic acid cleaving reagents, wherein, after coupling, the first adaptor-indexer strands are fully or partially single-stranded,

wherein the nucleic acid fragments to which offset adaptors and first adaptor-indexer strands have been coupled are binary sequence tags.

122. The method of claim 121 further comprising treating the nucleic acid sample to result in full or partial complementary sequences hybridized to the first adaptor-indexer strands.

123. The method of claim 122 wherein treatment of the nucleic acid sample to result in full or partial complementary sequences hybridized to the first adaptor-indexer strands is accomplished by hybridizing second adaptor-indexer strands to the coupled first adaptor-indexer strands.

124. The method of claim 123 further comprising covalently coupling the second adaptor-indexer strands to the nucleic acid fragments.

125. The method of claim 122 wherein treatment of the nucleic acid sample to result in full or partial complementary sequences hybridized to the first adaptor-

indexer strands is accomplished by filling in the single-stranded portion of the first adaptor-indexer strands.

126. A method of producing binary sequence tags from nucleic acid fragments in a nucleic acid sample, the method comprising

Sub A6

(a) incubating a nucleic acid sample with one or more first nucleic acid cleaving reagents to produce nucleic acid fragments,

(b) mixing one or more first offset adaptor strands with the nucleic acid sample and covalently coupling the first offset adaptor strands to the nucleic acid fragments, wherein, after coupling, the first offset adaptor strands are fully or partially single-stranded,

(c) treating the nucleic acid sample to result in full or partial complementary sequences hybridized to the first offset adaptor strands,

(d) incubating the nucleic acid sample with one or more second nucleic acid cleaving reagents to produce nucleic acid fragments with sticky ends, wherein the second nucleic acid cleaving reagents cleave at a site offset from their recognition sequence, wherein each first offset adaptor strand has a recognition sequence for at least one of the second nucleic acid cleaving reagents,

(e) mixing one or more first adaptor-indexer strands with the nucleic acid sample and covalently coupling the first adaptor-indexer strands to the nucleic acid fragments, wherein each first adaptor-indexer strand has a different end sequence, wherein each end sequence of the first adaptor-indexer strands is compatible with a sticky end generated by the second nucleic acid cleaving reagents, wherein, after coupling, the first adaptor-indexer strands are fully or partially single-stranded,

wherein the nucleic acid fragments to which offset adaptors and adaptor-indexers have been coupled are binary sequence tags.

Add A7